DOCUMENT RESUME

ED 204 413 TM 810 487

AUTHOR Stiggins, Richard J.
TITLE A Comparison of Direct and Indirect Writing Assessment Methods.
INSTITUTION Northwest Regional Education Lab., Portland, Oreg. Clearinghouse for Applied Performance Testing.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE Apr 81
GRANT 400-80-0105
NOTE 16p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis: Elementary Secondary Education: *Evaluation Methods: Higher Education: Objective Tests: Scoring: Testing: *Test Reliability: *Test Validity: *Writing Evaluation: Writing Skills
IDENTIFIERS Writing Samples

ABSTRACT

An area of current concern is that of the advantages and disadvantages of measuring writing proficiency directly via writing samples, and indirectly via objective tests. Much research has been completed documenting the correlation between direct and indirect measures. However, there had not yet been a systematic and detailed conceptual analysis and comparison of the two approaches. Such an analysis is presented in this paper. Direct and indirect writing assessment strategies are compared and contrasted in terms of the relationship each has to specific classroom decision-making situations, the components of writing assessed, practical testing matters such as user attitudes and testing costs, characteristics of test exercises, examinee response factors, test scoring and reporting procedures, and procedures for determining test quality. Conclusions are drawn regarding contexts when each approach may be useful. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204413

A COMPARISON OF DIRECT AND INDIRECT WRITING ASSESSMENT METHODS

Richard J. Stiggins

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM810487

Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

April 1981

The material presented herein was developed by the Clearinghouse for Applied Performance Testing (CAPT) under grant #400-80-0105 with the National Institute of Education (NIE) of the Department of Education. The opinions expressed in this publication do not necessarily reflect the position of NIE and no official endorsement by the Institute should be inferred.

ABSTRACT

There has been a great deal of discussion recently regarding the advantages and disadvantages of measuring writing proficiency directly via writing samples and indirectly via objective tests. Further, a good deal of research has been completed documenting the correlation between direct and indirect measures. However, there has not yet been a systematic and detailed conceptual analysis and comparison of the two approaches. This paper presents such an analysis. Direct and indirect writing assessment strategies are compared and contrasted in terms of the relationship each has to specific classroom decision making situations, the components of writing assessed, practical testing matters such as user attitudes and testing costs, characteristics of test exercises, examinee response factors, test scoring and reporting procedures, and procedures for determining test quality. Conclusions are drawn regarding contexts when each approach may be useful.

A COMPARISON OF DIRECT AND INDIRECT WRITING ASSESSMENT METHODS

There are two viable approaches to the assessment of writing proficiency. One is the direct method. It relies on actual samples of student writing to judge writing proficiency. The second is the indirect method, which relies on objective tests. Research on the correlation between the two approaches reveals a consistent and relatively strong relationship at various educational levels. Listed below are six studies that correlated objective language usage test scores with scores obtained on writing sample-based assessments.

Researcher(s)	Date	Subjects	N	Correlation
Breland, Colon & Rogosa	1976	College freshmen	96	.42
Breland & Gaynor	1979	College freshmen	819	.63
			895	.63
			517	.58
Huntley, Schmeiser & Stiggins	1979	College students	50	.43-.67
Godshalk, Swineford & Coffman	1966	High school students	646	.46-.75
Hogan & Mishler	1980	Third graders	140	.68
		Eighth graders	160	.65
Moss, Cole & Khampalikit	1981	Fourth graders	84	.72-.76
		Seventh graders	45	.60-.67
		Tenth graders	98	.20-.68

The results of these studies suggest that the two approaches assess at least some of the same performance factors, while at the same time each deals with some unique aspects of writing skill. This paper presents a detailed analysis of these common and unique dimensions and compares direct and indirect assessment methods on the basis of seven specific criteria:

1. The impact of each method on the various educational decisions teachers have to make
2. The components of writing assessed by each method
3. Practical considerations in testing, such as user attitudes, test acquisition options and testing costs
4. Characteristics of the test exercises
5. Factors related to the examinee's response to the exercises
6. Procedures used to score the test and report the results
7. Procedures used to evaluate test quality

The comparison of direct and indirect methods concludes with a summary of the major advantages and disadvantages of each, and a review of the specific roles that each plays in educational decisions.

Educational Context and Decision Making

Although fundamentally different in form, both direct and indirect writing assessment methods can be useful in educational assessment. Each provides a slightly different kind of information regarding a student's ability to use or recognize standard written English.

In direct assessment, the examinee must actually write in response to a given prompt; results are then evaluated according to prespecified criteria. In indirect assessment, the examinee is asked to judge the appropriate use of language in a series of objective test items which often follow a multiple choice format. Actual writing is not required. Each testing method requires that the examinee apply previously acquired knowledge about language usage.

Each approach provides information that is useful in making a variety of educational decisions--including those which involve instructional management, selection and evaluation.

Instructional management decisions, which include diagnosing student strengths and weaknesses, placing students in proper writing programs, and helping students make vocational and educational choices, can be based on either measure. Direct measures of writing proficiency are valuable in this context so long as resources are available to conduct relatively detailed analyses of results. Indirect writing assessment methods will also serve well if criterion referenced tests are used. Such tests break students' overall writing performance into component parts, allowing for a detailed analysis of skill development.

Both indirect and direct methods are suitable for selecting examinees for admission to special programs, or certifying minimum competencies. In the case of direct assessment, the scoring criteria used to rate writing samples must explicitly cover predefined essential skills. Similarly, with indirect measures, test items must test those skills essential for certification of competence.

Program planning (or evaluation) can also be based on direct or indirect assessment. In this case, the scoring criteria (for direct measures) or the test items, (for indirect) must reflect intended program outcomes. Or to put it another way, whichever assessment method is used must test that which is taught.

Assessment Focus

Direct and indirect assessments focus on different components of writing. Direct assessment measures actual composition skill. Indirect tests ability to use--or recognize proper use of--the conventions of effective writing: grammar, punctuation, sentence construction, organization, and so on. Direct assessment provides necessary and

sufficient information for drawing conclusions regarding a student's writing proficiency. Indirect assessment, on the other hand, provides necessary--but not always sufficient--information for evaluating a student's writing proficiency.

A review of the kinds of traits measured in the two approaches reveals that indirect assessment tends to cover highly explicit constructs in which there are definite right and wrong responses (e.g., grammar is either correct or it is not). Direct assessment, on the other hand, tends to measure less tangible skills (e.g., persuasiveness), for which the concept of right and wrong is less relevant.

Practical Testing Considerations

Consideration of several important practical factors is essential in ensuring a quality assessment. These factors reveal additional differences between direct and indirect assessment methods.

Key Attitudes. Users' attitudes are vital. With direct assessment, assessors and users of the test results must be willing to invest time, money and effort to conduct a writing assessment that calls for complex testing procedures (outlined below). In the case of indirect assessment, users must be willing to accept a proxy measure; that is, a test that covers component skills of writing without actually requiring students to write. Given the appropriate attitudes, either direct or indirect assessment will most probably have its desired impact. Otherwise, problems can be anticipated.

Test Acquisition and Development. In either direct or indirect assessment, the examiner has two choices: (a) selecting an already existing test or (b) constructing a new test.

If one wishes to use previously developed exercises (and scoring criteria for direct assessment), then what's needed to plan the assessment includes (1) technical expertise in writing, to specify which writing skills will be assessed; (2) test evaluation skills to investigate available options and select test items that measure the skills to be assessed; and (3) organizational skills to set up, administer, score and report the results of the assessment.

Developing a new direct assessment instrument, which involves creating a new set of exercises and criteria for scoring, also demands organizational skills and technical writing expertise. In addition, however, someone with psychometric expertise will be required to evaluate the validity and reliability of the assessment procedures, and refine exercises and criteria as necessary.

Developing a new indirect assessment instrument requires (1) technical expertise in writing, to plan the assessment; (2) skill in item writing or selection, to construct the new test; (3) organizational skills to pilot test, analyze and select the new items; and (4) psychometric expertise to evaluate the test's reliability and the validity.

In short, developing new instruments for either testing approach requires substantially more expertise and staff time than does use of existing assessment instruments.

Testing Costs. Cost is perhaps the single most important practical consideration in deciding what assessment approach to use. There are three kinds of costs to be considered: developmental costs, test administration costs and test scoring costs. The cost factors for direct or indirect assessment vary considerably and depend, to a great extent, on whether a new test is to be developed, or an existing test is used. The following lists indicate the most important factors affecting developmental cost in each of four contexts:

Developmental Costs/Previously Developed Direct Assessment

1. Staff time required to plan the assessment
2. Staff time required to secure, review, evaluate and select existing exercises and a scoring guide
3. Cost of producing all necessary test materials

Developmental Costs/New Direct Assessment

1. Staff time associated with planning the assessment
2. Time required to develop exercises
3. Time required to develop a scoring guide
4. Time and administrative costs associated with field testing the exercises and scoring procedures
5. Costs of producing all necessary test materials

Developmental Costs/Previously Developed Indirect Assessment

1. Staff time associated with planning the assessment
2. Staff time necessary to review, evaluate and select the test
3. Cost of purchasing all necessary test materials

Developmental Costs/New Indirect Assessment

1. Time required to plan the assessment
2. Costs associated with item writing (or selection from an item pool, if available)
3. Time to pre-test items, analyze the data and assemble the test
4. Cost associated with norming the test (if necessary)
5. Cost associated with the production of all necessary test materials

The test administration costs will generally remain constant regardless of whether the assessment is direct or indirect, or whether new or existing test items are used. Administrative costs are a function of the time required to plan the test administration, select and train test administrators, coordinate the materials distribution, administer the test, and coordinate the collection of test materials.

Test scoring costs differ substantially between the two methods. Scoring costs for direct assessment are determined by at least five factors:

1. Time associated with planning relatively complex scoring procedures
2. Selecting, recruiting and training scorers
3. Staff time required to read and rate the writing samples
4. Evaluation of the reliability of the scores
5. Time for processing the scores for reporting

Scoring indirect assessment results is less complex, and therefore generally less time consuming and costly. It involves either computer time to machine score tests, or personnel time to hand score. If a new indirect instrument is used, there is the added cost of evaluating its reliability. As with direct assessment, there are also costs associated with processing the scores for reporting.

If security is not breached, or if test security is not an issue, both direct and indirect test exercises can be reused. However, the impact of test reuse on overall testing costs depends on which of the two approaches is followed. With direct assessment, test development costs are low, while scoring costs are high. The opposite is true of indirect tests: development costs are high, while scoring costs are low. Reuse clearly minimizes test development costs, and therefore, can significantly reduce the cost of indirect assessment. Reuse will not substantially lower the cost of direct assessment, however, because it has no impact on scoring, which must recur with every new administration. Before deciding which approach to use, it is wise to conduct a cost/benefit analysis that takes all factors--including possible test reuse--into account.

Summary. Attitude is an important consideration in selecting a testing approach. With direct assessment, what counts is willingness to invest time, effort and money in a complex assessment process; with indirect assessment, one must be willing to accept a proxy measure. In either case, it is possible to select an existing assessment instrument, or develop a new one. Either decision has implications for the kind of expertise required to conduct the assessment, and for the costs incurred. The principal costs associated with direct assessment relate to scoring, while the principal costs associated with indirect assessment relate to test development.

Characteristics of Test Exercises

There are some fundamental differences in the kinds of test exercises used in direct and indirect writing assessment. First, the exercises differ in form. Direct assessment exercises generally take the form of a sentence or short paragraph that invites the examinee to respond to a question, state an opinion, resolve an issue, explain a process, recount an event, or simply express his/her feelings. The exercise, if well constructed, identifies for the examinee the form of writing to be produced, the audience to be addressed, and the purpose for the writing. Indirect assessment items frequently follow a multiple choice format though fill-in questions are sometimes used. Various interlinear forms, as well as sentence combining items are also common.

As a result of differences in format, direct assessment exercises are considerably more flexible than indirect. With direct assessment, the stimulus can be auditory or visual and can be quite true to life (e.g., writing a job application letter). Indirect test items, on the other hand, are generally constrained by the multiple choice (or other) format. Therefore, while direct assessment exercises can be made to closely approximate "real world" writing, objective test items are somewhat more artificial.

The manner in which the examiner exercises control over the skills tested differs too. In the case of the objective test, the examiner controls the kinds of skills tested by selecting test items that relate to those skills. Careful construction and selection of items can give the examiner very precise control over the specific skills tested. In direct assessment the examiner has some degree of control over the kinds of skills tested by selecting or developing writing exercises that specify the form of writing to be demonstrated (e.g., essay, letter, narrative), the audience to be addressed and the purpose for the writing. But the degree of control is not as great as with an objective test.

This difference can give rise to validity problems with direct assessment if the test is not developed and used carefully. Under some circumstances, the use of a writing sample to judge certain skills leaves the examiner without assurance that those skills will actually be tested. For example, consider mechanics. Less-than-proficient examinees composing essays might simply avoid unfamiliar or difficult sentence constructions. But if their writing contained no obvious errors, an examiner might erroneously conclude that they were competence, when, in fact, their skill in mechanics had never been fully tested.

With indirect assessment, sampling error is controlled by forcing the examinee to demonstrate mastery or nonmastery of specific elements of writing. Examinees cannot construct a response to suit themselves; they must respond within the framework of the test format.

Consequently, with regard to test exercises, one maximizes the authenticity and flexibility by using direct assessment, but in doing so, may sacrifice some control over the kinds of skills tested. With

interpretation is possible. That is, examinees can be compared to each other or to a common standard of acceptable performance.

Judging Test Quality

The factors commonly considered in judging the psychometric adequacy of a test are reliability and validity. The relationship between direct and indirect writing measures is discussed in this section in terms of these two factors.

Reliability. Like any test, a writing test must yield dependable or reliable scores to be useful; in both writing assessment approaches, reliability takes various forms. Test scores must be stable over time, across parallel test forms, across different parts of the same test and across raters. Otherwise, the examiner will not know if a score accurately reflects an examinee's proficiency.

The following brief examples illustrate how poor reliability can affect test results in four ways. Suppose a direct writing assessment were administered to the same students twice, the second administration following a two- to three-week interval. And suppose that even though no writing instruction had taken place, the scores obtained the second time were totally different from those achieved the first time--for nearly every examinee. The examiner would not know which score (if either) to depend on as the true reflection of the student's proficiency. Or, suppose two writing exercises were developed to measure exactly the same skills and yet when both were administered to a student, the exercises resulted in totally different estimates of proficiency. Again, the examiner would not know which score was the better indicator of proficiency. Or what if ostensibly equivalent forms of a performance test resulted in totally different estimates of proficiency. The examiner would not know which form to rely on. Or, from a fourth perspective, suppose two judges evaluated the same performance and drew totally different conclusions regarding proficiency. In this case, the examiner would not know which judge to rely on.

When scores on tests are unstable over time, differ considerably across ostensibly equivalent exercises or test forms and/or differ substantially across evaluations of proficiency by independent judges, there is reason to question the usefulness of the assessment procedures. When these differences occur, it is possible that the examinee's score was influenced by administration time, the particular exercise or test used, or the relative qualifications of the rater who happened to evaluate the response--all factors independent of the examinee's real proficiency. These independent (and irrelevant) factors are undesirable determinants of test scores. Only when writing assessment procedures yield scores that are stable over time, across exercises, test forms and independent evaluators, can those scores be confidently used for educational decision making.

Consistency across raters is not an issue in the case of objective instruments. Consistency over time, across items and across test forms, however, remains relevant for both assessment approaches.

One way of safeguarding test reliability is to be aware of potential sources of unreliability. These factors differ somewhat according to assessment approach. In direct assessment, inaccurate scores can arise from (1) poor exercises (e.g., ambiguity, bias), (2) poor test administration procedures or environment, or (3) poor scoring procedures (e.g., rater leniency, halo effects, tendency toward middle rating). Poor test items and test administration procedures can also be problems with indirect assessment. Poor scoring procedures are not a factor in indirect assessment. However, on the other hand, the format of objective tests often encourages guessing--a potential source of error.

One dimension of score reliability that becomes particularly important in large scale writing assessments is score scale consistency across time (say, from one year to the next). Some assurance is needed that scores from one year are directly comparable to scores on the same test the next year, particularly if a high school diploma or other certificate of achievement rests in the balance. In the case of indirect assessment, this equivalence is accomplished via complex but well developed statistical procedures. In direct assessment, lquivalence is attained by carefully training raters to be sure that they understand and apply the same rating criteria each year. In addition, criteria must be sufficiently explicit to minimize variability among raters.

Validity. Two important considerations relate to the validity of writing tests. The first is content validity, or the extent to which objective test items or writing exercises and scoring procedures do, in fact, represent the kinds of skills that are the intended focus of the test. In both direct and indirect measures, this kind of validity is a matter of expert judgment: Test items or exercises and scoring guides (whichever are used) must be reviewed by independent writing specialists to ensure their appropriateness.

The second type--criterion related validity--must be considered from two perspectives: internal and external. Internal validity is verified by correlating writing test scores with other simultaneously administered measures of writing proficiency. In the case of indirect measures, internal validity must always be verified by showing a high correlation between the objective test score and a score attained through the use of actual writing samples (direct measures). Internal validity of direct measures can be verified by correlating scores with other writing performance indicators, such as other writing samples or course grades.

The external validity of a writing test reflects the extent to which the scores correlate with other valid indicators of writing proficiency. For example, scores on either direct or indirect measures are valid from an external perspective if they predict subsequent English course grades.

Summary. Reliability and validity considerations for direct and indirect measures are quite similar. In the case of direct measures, score stability is important over time, across exercises, across test forms and across raters. Consistency across raters is not an issue with indirect measures, however. In both cases, sources of inaccurate scores include poor test items and improper test administration. Unique sources of score inaccuracy include guessing in indirect measures and

poor or inconsistent scoring in direct measures. Score scale equivalance from one administration to the next is maintained via statistical methods for indirect measures, and via rater training for direct. Content validity is relevant to both types of writing tests and is verified in both via expert judgment. Criterion related validity, also important in both cases, is verified in terms of correlations with other indicators of writing proficiency.

Conclusion: Comparing Assessment Options

Direct and indirect approaches to writing assessment are perhaps best compared in terms of their relative advantages and disadvantages, and the primary ways in which each can use used.

Advantages and Disadvantages. The major advantages of the direct assessment option are (1) the extent of information provided about examinees' writing proficiency; (2) high fidelity of the stimulus and response; (3) the adaptability of exercises to a variety of relevant real world writing circumstances; (4) high face validity of writing samples; and (5) relatively low test development costs.

The major advantages associated with the indirect assessment are; (1) high score reliability, (2) relatively low test scoring costs, and (3) high degree of control over the nature of the skills tested.

The disadvantages of the direct method include; (1) high scoring costs, and (2) the potential lack of uniformity of proficiencies assessed among examinees.

The disadvantages of the indirect method are; (1) lack of fidelity to real world writing tasks, (2) heavy reliance on examinees' reading rather than writing proficiency, and in many cases (3) lack of face validity in the objective measure.

Uses of Writing Tests. Given these advantages and disadvantages, we can identify the conditions under which direct and indirect assessments can be used to make the kinds of educational decisions outlined in the introduction. For example, in instructional management, both direct and indirect writing assessments can play a significant role in the diagnosis of strengths and weaknesses, course placement, and guidance decisions.

If used for diagnosis, direct assessment is best with relatively small groups, so that a teacher can conduct a detailed analysis of each student's writing. Direct assessment may be less useful for diagnosis with larger groups because analytical scoring is time consuming, and consequently costly.

If indirect writing tests are criterion referenced, they can also be used for diagnosis of strengths and weaknesses. The tests must include items specifically referenced to important elements of writing and must yield scores that reflect a student's mastery of each element. If these conditions are satisfied, an objective test can be quite diagnostic.

Direct assessment can play a role in course placement and guidance if relatively low-cost scoring procedures are selected. In most cases, this means using holistic scoring procedures to obtain an overall estimate of each student's performance as it relates to that of other students in the group. This kind of comparative test score can aid both students and teachers in educational and vocational planning.

Indirect assessment is more commonly used in placement and guidance than is direct assessment. It is not uncommon to see a published, norm referenced test of language usage proficiency used to rank students' usage skills for placement into various developmental writing courses. Such tests are often used to help students decide whether to enter college programs or pursue careers in which writing proficiency is prerequisite to success.

Writing tests can also play a role in the selection of examinees; for instance, in admitting students to a remedial or advanced program or in certifying minimum competencies. Direct assessment can be used for selection, and is, in fact, the best choice when writing proficiency is the sole or primary selection criterion. For example, when a test is used to determine which students will receive four-year scholarships for a college writing program, the test with maximum fidelity--a direct assessment measure--would be the test of choice. Indirect measures can also be used in selection, however, and are acceptable whenever writing proficiency is one of many selection criteria. With respect to certification, direct measures are useful whenever the scoring scheme has been so designed that students' performance is rated according to what are considered the minimum acceptable competencies. Indirect measures can be useful in this context if test items are carefully selected to represent minimum acceptable skills.

Writing tests can also play a role in program planning or evaluation decisions. For example, writing tests may be used in (1) survey assessment, (2) formative program evaluation, and (3) summative program evaluation. In survey assessments, such as statewide programs, tests are administered to a representative sample of students in order to generalize about the proficiency levels of the larger group. Since these surveys typically include large samples of students (often ten thousand or more), scoring costs are a factor. Holistic scoring, a highly cost-efficient approach, is often the choice with direct assessment. Some state programs take advantage of the scoring efficiency associated with indirect measures, either in place of or in combination with direct assessment.

Appropriate use of either method in formative or summative program evaluation requires that intended course outcomes be reflected in either the direct assessment scoring criteria or the indirect assessment items. In short, both direct and indirect assessment can be useful in evaluation provided there is close correlation between the tests and the instructional program.

This analysis of the similarities and differences shows how both direct and indirect assessment can play a valuable role in measuring student writing proficiency. Though one method is not inherently superior to the other, their relative appropriateness and usefulness varies

according to the educational assessment context and the decisions to be made. Anyone planning to assess writing proficiency would be wise to analyze the strengths and shortcomings of each option in light of the informational needs to be addressed.

REFERENCES

Breland, H.M., Conlon, G.C., and Rogosa, D. A preliminary study of the Test of Standard Written English. Princeton, NJ: Educational Testing Service, 1976.

Breland, H.M. and Gaynor, J.L. A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 1979, 16, 119-128.

Godshalk, F.I., Swineford, F., and Coffman, W.E. The measurement of writing ability. New York, NY: College Entrance Examination Board, 1966.

Hogan, T.P. and Mishler, C. Relationships between essay tests and objective tests of language skills for elementary school students. Journal of Educational Measurement, 1980, 17, 219-227.

Huntley, R.M., Schmeiser, C.B., and Stiggins, R.J. The assessment of rhetorical proficiency: The role of objective tests and writing samples. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco, CA, 1979.

Moss, Pamela A., Cole, Nancy S., and Khampalikit, Choosak. A comparison of procedures to assess written language skills in grades 10, 7 and 4. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA, 1981.